

Armed Services Technical Information Agency

AD

19361

NOTICE: WHEN GOVERNMENT OR OTHER DRAWINGS, SPECIFICATIONS OR OTHER DATA ARE USED FOR ANY PURPOSE OTHER THAN IN CONNECTION WITH A DEFINITELY RELATED GOVERNMENT PROCUREMENT OPERATION, THE U. S. GOVERNMENT THEREBY INCURS NO RESPONSIBILITY, NOR ANY OBLIGATION WHATSOEVER; AND THE FACT THAT THE GOVERNMENT MAY HAVE FORMULATED, FURNISHED, OR IN ANY WAY SUPPLIED THE SAID DRAWINGS, SPECIFICATIONS, OR OTHER DATA IS NOT TO BE REGARDED BY IMPLICATION OR OTHERWISE AS IN ANY MANNER LICENSING THE HOLDER OR ANY OTHER PERSON OR CORPORATION, OR CONVEYING ANY RIGHTS OR PERMISSION TO MANUFACTURE OR USE OR SELL ANY PATENTED INVENTION THAT MAY IN ANY WAY BE RELATED THERETO

Reproduced by
DOCUMENT SERVICE CENTER
KNOTT BUILDING, DAYTON, 2, OHIO

UNCLASSIFIED

19361

Department of Psychology
University of Missouri
Columbia

STUDENT BEHAVIOR RATINGS AS A CRITERION
OF TEACHER EFFECTIVENESS

by

Thomas C. Burgess

Technical Report No. 3 for Contract NONR 649(00)

between

University of Missouri and Office of Naval Research

August 1953

CONTENTS

	Page
PART I Theoretical Considerations in Development of Criteria of Teaching	
Effectiveness	1
The Ultimate Criterion	2
Substitute Criterion Measures.	3
An Extension of the "Pupil Gain" Criterion	5
Summary.	6
PART II An Exploratory Study of Student Behavior Ratings as a Criterion of Teacher Effectiveness	7
The Behavior Dimensions.	7
Development of the Observation Scales	8
Development of the Values Scales	10
The Rating Scales	10
Results	11
Conclusions and Recommendations.	12
REFERENCES	14
APPENDIX A Items, Scoring Weights, and Item Analysis.	
APPENDIX B <u>Observer Check List.</u>	
APPENDIX C <u>The Ideal College Classroom.</u>	
APPENDIX D <u>Missouri Studies on Teaching Effectiveness, Classroom Rating Scale.</u>	
APPENDIX E Data on Classes Observed.	

PART I

THEORETICAL CONSIDERATIONS IN DEVELOPMENT OF CRITERIA OF TEACHING EFFECTIVENESS

The notion that people should be carefully selected in terms of their abilities, aptitudes, interests, personality, etc. to perform certain kinds of jobs has been accepted rather generally by this time. Numerous instances of the use of selective devices might be cited reaching, indeed, as far back in history as the Biblical times when Gideon's famous Three Hundred were selected by a kind of performance test (Judges 7:1-8). Others may be mentioned which are as recent as the practices in selection of business executives, indicated in a survey by Fortune Magazine, which emphasize the evaluation of the candidate's wife (12,13).

The extent of the belief in selection procedures is indicated by the variety of situations and personnel to which they are applied. In the recent war the Army selected men for officer training by intelligence tests, recommendations by officers and others, and a group interview with the candidate. Many business concerns now use psychological tests and interviews to select many of their employees. Many professional and semi-professional groups have sought and obtained legal bases for selection programs to prevent the untrained, the unfit, and the charlatan from practicing their specialties. Such provisions apply to such diverse fields as nursing, law, embalming, teaching, psychology, real estate selling, and barbering. In many of these fields the presentation of a certificate of the completion of training is sufficient to obtain a license or a certificate. But in many others the candidate must also pass a set of examinations before being admitted to practice.

Experience and research have shown that frequently a selection program is ineffective if it is not based on research on such things as the characteristics necessary or conducive to success in the job, the precise nature of the job, and the effectiveness of certain tests or background information in predicting success. This is certainly the case in many of the occupations mentioned. In most of these cases the selection procedures have been set up on an a priori basis, using the subjective judgment (admittedly the best information available at the time) of men in the field. To a considerable extent, such judgments are still the best information that is available, and the evidence used to predict performance in the field cannot now be shown to be inferior to any other kinds of measures that might be proposed. Now, why is this the case?

An examination of the list of occupations given will indicate that these groups have two things in common: (1) They are all, in a sense, service occupations in which incompetent service will harm the public and reflect on the profession itself. Therefore the public should be protected, both for the sake of the public and for the sake of the profession. (2) They are all occupations in which it is more than usually difficult to specify exactly the nature of competence in the field, or in which this has not been done. These two facts combine on the one hand to produce a pressure to select practitioners carefully, and on the other hand to make it extremely difficult to find out whether the selection procedures used are doing a good job or not.

Research done in scientific selection programs has shown that it is possible to improve selection of workers in the clerical, manual, and mechanical fields of work. It should also be able to help improve selection in the professions. But the very existence of any scientific evaluation of a selection program depends upon the availability of some good measures of success on the job, some good acceptable criterion. It is just this criterion, or rather the lack of it, which has been holding up the effective evaluation of these professional selection programs. No evaluation leads to no improvement and to the impossibility of knowing whether a change in selection procedure is actually an improvement or not.

It will have been noted that teaching was included in the list of professions that use a legalized selection program. This certification, as it is usually called in education, is most often based on the completion of a prescribed course of study, which usually includes some actual practice in teaching under supervision of a "master" teacher. The effect of this kind of selection program is really to shift the responsibility for selection on to the teacher training institution. At this level

one finds an astonishing diversity of selection policies. These range from a low point of selection on the basis of freedom from gross, incapacitating emotional maladjustment and ability to pass college courses. At the other extreme we find some teacher training institutions setting high standards of selection in terms of academic achievement, emotional stability, experience in leadership, and other qualities that are presumed to be desirable in a teacher. But it is typical that graduates of both types of training program may obtain the same certificate, and in the same state. And it is worth repeating, that in both cases the college has set up its program on the basis of a maximum of intuition and a minimum of information.

At this point it may be pertinent to digress for a moment in order to point out that the problem of recruitment is properly considered a part of the problem of selection. The sources of applicants and the kind of appeal that brings them to the point of application often determine the kind of applicants that the college or profession has from which to select its members. Should one not ask what kind of appeal can be made for a profession which offers high rewards of an intangible nature combined with material rewards on a bare subsistence level? Where will the applicants for such a position come from? And of what quality will they be? Any college must consider this problem in setting up its selection standards, because a college without students finds itself in an embarrassing position.

The problem of improving the selection of teachers has occupied the attention of a number of psychologists and educators for many years. While the results of this concern have been disappointing, they have no need to apologize for spending their time on the problem. The problem is of crucial importance. The teacher affects the lives of nearly all. And in these complex times the nation cannot afford to squander any of its great human potential any more than it can afford to squander its natural resources. Unlike natural resources, which may be conserved by using them sparingly, the nation's human potential can be conserved only by developing it to its utmost. And it is apparent that this development of the human potential is the primary responsibility of the teacher. It is, therefore, imperative to find ways of selecting and training teachers who can and will promote this development. Such teachers are needed in the public schools. They are also needed for the training programs in the armed forces, where survival of the individual and of the nation is the test of their effectiveness. Here the results often become apparent all too soon. The problem of teacher selection is therefore well worth attention.

The Ultimate Criterion

As stated earlier as a general principle, the crux of a selection program is the criterion. Obtaining the criterion for any selection program is a headache, but finding a satisfactory criterion for the success of a teacher is a headache of epic proportions. Why? Because the teacher's product is a person. The teacher tries to prepare his students to lead rich, effective, and satisfying lives. So the ultimate criterion of the teacher's success is the richness, the effectiveness, and the satisfyingness of the later life of the student (5). And just how can the richness, effectiveness, etc. of a person's life be measured? It can't be done, at least right now. But if something less than this is used, such as total income in dollars and cents, the teacher objects that he is being judged unfairly and not in terms of his total objectives. And he is right.

To further complicate the problem, the teacher is only one of many teachers who have contributed to the development of the student during his school career. To judge the teacher fairly, his influence on the student must be separated from the influence exercised by the other school personnel who have helped to make him what he is. This is a tough problem and would take a long and immensely complicated study to solve. But if appropriate criterion measures could be developed, this problem could in its turn be taken care of by means of multivariate analysis.

But here again the teacher objects that the school is not wholly responsible for the development of these people. Rather, they share the responsibility with the community, the parents, the church, etc. After all, children are actually in school only a small portion of the time--and they are already well along in their development before the school ever sees them. This must be recognized and dealt with, for it is indeed true. At this point it may be seen that, theoretically at

least, this problem can be handled in the same way as the preceding one. This is perhaps possible. But in view of the practical difficulties involved, it seems unlikely--unless another way could be found.

Substitute Criterion Measures

There are a number of other ways of trying to measure teacher success that have been tried in the past. These ways might be summarized as being of three general types: (1) the measurement of changes in pupil behavior, (2) the measurement of teacher behavior, and (3) the measurement of teacher characteristics. All of these attempts to solve the criterion problem have several things in common: (1) they can be measured or estimated easily, (2) they can be obtained in a relatively short time, (3) they can be related to the work of the particular teacher whose work is being evaluated, and (4) educators are dissatisfied with all of them. To understand this dissatisfaction, it is necessary to evaluate each of these methods in terms of the characteristics which are needed in any such substitute criterion measure.

What are these needed characteristics? The first and most important requirement for a substitute criterion measure is relevancy or validity. That is, the measure to be used must be related to the ultimate criterion. This requirement immediately arouses a problem. It has already been shown that the ultimate criterion of education is elusive, perhaps hopelessly so. How, then, can it be known whether or not any substitute criterion measure is related to the ultimate criterion? The answer is that it cannot be known; it must be assumed. As Thorndike has said, "In practice, the complete ultimate criterion is never available The result is, as indicated earlier, that the relevance of a particular criterion measure usually must be estimated very largely on rational grounds with only limited help from empirical data." (11, p. 125.) One is thus forced to assume, on rational grounds, the relevance of any criterion measure of which use is to be made. There is, furthermore, no escape. This problem of making a decision about the relevance of our criterion measure cannot be avoided since, to quote Thorndike, "Relevance is the absolutely fundamental requirement of a criterion measure." (11, p. 125.)

The remaining requirements of a substitute criterion measure are that it be to some extent reliable and practical. Nothing further need be said about these latter two requirements, since all of the suggested criterion measures appear to either meet these requirements or could be refined to the point at which they could satisfy them.

It is then necessary to examine the three types of criterion measures that have been used and to see how well they satisfy the requirement of relevance.

The pupil gain criterion. The first type of criterion measure is the so-called "pupil gain" criterion. This involves the use of tests, usually, to determine how much the pupil has learned during the period of his exposure to the teacher in question. The assumption that seems to be involved in the use of this type of criterion measure is that what the student learns now is related to what the student will do later in life. This seems to be a reasonable assumption to make. Most people are, therefore, inclined to accept this type of criterion measure as being one which is adequate.

But, as has been previously indicated, educators seem to be dissatisfied with this criterion. Why? It appears that their objections are based on the present limitations of this approach. There exist tests to measure pupil gains in subject-matter knowledge, understanding of subject matter, study skills, and related achievements. No satisfactory instruments have been developed, however, to measure pupil progress toward other immediate educational goals such as the achievement of independence, the development of ability to work together cooperatively and harmoniously in the solution of common problems, the development of the ability to draw together information and experiences from diverse sources to invent a solution to a present or anticipated problem, etc. Insofar as these types of objectives are important to a school or a teacher, the "pupil gain" criterion must be considered as only a partial substitute criterion. That is, a teacher who considers these intangible goals to be important, objects to being evaluated on the basis of subject

matter achievement alone. In other words, the educator admits the relevance of this type of criterion but insists that right now the available techniques do not permit a sufficiently broad coverage of educational objectives.

The measurement of teacher behavior. The type of substitute criterion measure that attempts to use teacher behavior is usually approached by some method of rating the teacher's behavior (or method). The ratings may be made by the teacher's supervisor, the students, or by visiting observers. The use of this type of measure as a criterion of teaching effectiveness involves the assumption that what the student learns now is related to what the student will do later in life, and that what the teacher does now is related to what the student learns now. As indicated in the discussion of the "pupil gain" criterion, the first part of this assumption seems to be justified. It also seems apparent that the last part of the assumption is equally justified. But on further examination it appears that the second part of this assumption does not take adequate cognizance of the multiplicity and complexity of the factors determining what the students learn in the classroom. That is to say, as we get further away from the actual behavior of the student, the factors which must be considered become more numerous as well as more complex in their interrelations. For example, it is known that the student's personality is a factor in the determination of what he learns from a given teacher in a given situation (14). It is also known that the behavior learned in the classroom is a function of what the teacher does (1, 2, 3). There is further suggestion that student behavior is also affected by teacher personality (4, 6, 9). In addition to these, it is asserted that all of these factors are in turn affected by certain aspects of the particular situation in which the teaching is done. In other words, to predict student outcomes would require a combination (the nature of which has yet to be demonstrated) of teacher personality (specific traits unknown), teacher behaviors or methods (unknown), and the special situation in which the teaching is done (the significant variables in this situation are also unknown). It is further apparent that in order to specify these various factors and their combination, it will be necessary to do a great deal of research utilizing a good, justifiable criterion of effectiveness.

None of what has just been said is particularly new. It has been said before in one way or another (5, 8, 10). Perhaps it has even been said by those who have used this type of criterion measure. But say it or not, they have used the teacher behavior criterion in a way which implies that the relationship between it and student behavior is a simple one. The evidence is clear. It is utter folly to continue to assume a simple relationship between teacher behaviors and student behavior. It is then apparent that it is equally futile to continue to assume any close relationship between teacher behaviors and the ultimate criterion of education.

The measurement of teacher characteristics. The third kind of measure which has been used as a criterion of teacher effectiveness tends to focus attention on the kind of person the teacher is. It would include ratings on teacher characteristics such as appearance, voice, poise, etc. as well as scores on tests designed to measure traits which are presumed to be desirable in teachers. This latter category would include personality test scores, intelligence test scores, and achievement test scores of the teacher.

The assumption involved in the use of these measures as criteria is quite similar to that made in the use of teacher behavior and could be stated as: what the student learns now is related to what the student will do later in life, and what the student learns now is related to the kind of person the teacher is. This assumption seems to be justified, but it is subject to the same criticisms made of the teacher-behavior type of criterion measure. In fact, it will be recalled that in making those criticisms, teacher behavior and teacher personality were listed as parallel factors in the prediction of student outcomes. It can be seen, therefore, that this type of criterion measure suffers from the same lack of relevance as the teacher behavior measures. Like the latter measures the teacher characteristic measure must be rejected as an unsatisfactory criterion.

In summary it may be said that all three types of substitute criterion measures are probably satisfactory with respect to the requirements of reliability and practicality. It also seems reasonable to conclude that a broadened "pupil-gain" type of criterion has a satisfactory degree of

relevance to the ultimate criterion. This cannot be said, however, of the "teacher behavior" and "teacher personality" criteria, which seem to be lacking in the degree of relevance needed for the productive investigation or evaluation of teaching.

The result of this analysis of existing substitute criterion measures is to leave, for the present, only one acceptable criterion: the "pupil-gain" type of measure. Since this is a severely limited type of measure at the present time, it would appear that psychologists and educators must turn their efforts to the extension of this type of approach so that it will encompass more of the educational objectives of present-day schools. In the following section it will be suggested that the "pupil gain" criterion be broadened to include estimates of other pupil classroom behaviors, in order to make the "pupil gain" criterion cover these intangible educational objectives.

An Extension of the "Pupil Gain" Criterion

As one visits classrooms he begins to notice that they are quite different. Some are active and cheerful; some are quiet and depressed; some are active and hostile; some are quiet and rebellious, etc. As he continues to observe classes, he notes that these characteristics seem to be independent of subject matter, educational level, general teaching methods, lesson plans, physical facilities, etc. That is, these differences seem to be unrelated to the more obvious and objective things to be seen in a casual inspection of the class. It may be assumed that these differences are associated with the teacher for the most part (teacher personality, methods, and situation remaining relatively constant). That is, one would expect to find less difference in one teacher's class from year to year than he would find between the classes of different teachers in any one year (assuming that the classes are visited not too soon after the beginning of the term or year). An indication of this is contained in the work of Anderson and Brewer (1, 2, 3). This would mean that the student behavior seen in the classroom can be attributed to the teacher.

But if this be true, then it would seem that these students are learning some sets of behaviors in school as a result, primarily, of exposure to this teacher. The fact that these behaviors are limited to the classroom and are felt by the students to be appropriate only to this classroom is a possible limitation, but perhaps not too serious a one. For these students are learning a set of behaviors that may very well be used by them in other social situations. These behaviors are being added to the repertory of behaviors that the student will have at his command to rely upon in future social situations of different sorts. It would seem logical to assume that there is some relationship between the behaviors the student learns and the effectiveness of his adjustments to life. Therefore, if these student behaviors can be specified and measured in some way, the result would be a set of criterion measures that could easily be justified. Such a set of criterion measures could be classified as a kind of "pupil gain" criterion. That is, since people learn to do what they are doing, the students can be assumed to be learning the behaviors they are exhibiting in the classroom. Under this assumption it is no longer necessary to pay especial attention to changes over a period of time, or to indicate how well they are learned. Rather the direction which the learning is taking would be of primary interest. Observe the behaviors being "practiced" in the classroom, locate those behaviors on some dimension of behaviors, apply values to the dimension (i.e., specify the part or region of the dimension which corresponds to the objectives of the school), and the result is a criterion measure.

So procedures may be set up to measure the ways in which student behavior differs in different classrooms. These procedures may require a somewhat different approach to the idea of measurement than that usually encountered. Here the measurement procedures would function more like a compass to indicate the direction of progress than like the usual measuring device, which functions like a thermometer to indicate the degree of progress in one specified direction. Thus classrooms might be evaluated in terms of their placement on a continuum of behavior extending from, say, the dependency of students on authority at one extreme to the exercise of initiative, independence, or self-direction in the solution of problems at the other. Assuming, as has been done above, that students are learning whatever it is they are doing, it becomes possible to say, after using this "compass," in what direction learning is proceeding. Just what kinds of behaviors are being learned in the classroom could be specified by using several of these "compasses."

This procedure would work in this way with respect to the dependency dimension. First, one would observe the behavior of students in the classroom. These behaviors would then be classified, according to their meanings (in this case for dependency) summarized and located somewhere on the continuum from Dependency to Self-determination. This location would then be compared with the location on the continuum of the "ideal," or most highly valued point based upon the educational objective. Such a comparison would yield a difference score for that class. This score would indicate how close this class is to practicing what is considered the ideal kind of dependency behavior. This would be a score which would indicate how close this class is to the educational objectives dealing with the dependency of behavior; hence it would be a score that is what is needed for a criterion measure of this type of behavior.

Once knowing the kinds of behaviors the students are learning in a class and having evaluated them in terms of what is known about the effects of these kinds of behavior or in terms of the values attached to such behaviors, an evaluative statement can be made about the class and its teacher. It can then be said that the teacher is "good" or that he is "bad." And when such statements can be made, a criterion measure has been produced that can and should be used in the development of scientific selection procedures.

This approach appears to have several advantages:

- (1) It yields some immediately obtainable measures that can be used for immediate research on the predictors of teaching effectiveness.
- (2) It avoids the necessity of considering the effects of the parents and the community--provided that assumptions about the consistency of classroom behaviors associated with particular teachers are borne out (it could be tested easily).
- (3) It avoids the necessity of considering the effects of other teachers--under the same assumptions.
- (4) It is felt that enough knowledge of psychology is now available to make it defensible to assume some of the necessary relationships between these variables and the ultimate criterion. It is more defensible in this respect than other criteria that have been used in the past which have assumed a relationship between the ultimate criteria and certain teacher behavior or teacher characteristic variables.
- (5) It provides for the separation of the variables on which ratings are made from the values that are attached to the various kinds of learnings that may occur. This makes the device more flexible and usable by persons having different values attached to the student behaviors involved. It also, incidentally, compels the formation of values by directing the attention of educators to the presence of variables that are often ignored or overlooked as intangibles.

Summary

1. The difficulties involved in the selection of teachers are due to the lack of suitable criterion measures.
2. Suitable criterion measures are lacking because of the inaccessability of the ultimate criteria, and the limitations, remoteness or inappropriateness of substitute criteria.
3. The observation, classification, and evaluation of student behavior in the classroom was suggested as an appropriate and promising criterion measure under the assumption that students are learning the behaviors they practice.

PART II

AN EXPLORATORY STUDY OF STUDENT BEHAVIOR RATINGS AS A CRITERION OF TEACHER EFFECTIVENESS*

This is the report of an exploratory study of the use of criterion instruments of the type suggested in Part I and in a previous report (7). The principal aim of the study was to devise and try out an instrument for the use of classroom observers, which would sample student behavior in three different behavior dimensions which might be considered to be related to important objectives of education.

As previously indicated (7), the scales to be developed should have the following characteristics: (1) emphasis on student behavior, (2) specific behavior descriptions, (3) definite provision for the application of values to the scales, and (4) a format which would not be too time-consuming. It was felt that the emphasis on student behavior would provide measures which would be highly acceptable as criterion measures, following the logic of Part I. The use of specific behavior descriptions should lead to scales which show a greater degree of agreement between observers using the scales to report on the same class. Making some specific provisions for the application of values to the scales should have the effect of making the scales usable by a larger number of institutions with differing values and objectives. The fourth requirement was stated as a desirable condition leading to economy of time and effort in applying the scales.

The Behavior Dimensions

There probably exist a relatively large number of different kinds of behavior that might be observed in the classroom and which would have some significance for the ultimate objectives of the educational process. Many different behavior dimensions might have been selected and used in the present study. For the purposes of this exploratory investigation, however, it was felt desirable to limit this number to the three dimensions which have been labeled Integration, Dependency, and Tension.

Integration. This dimension is concerned with the extent to which the students use a variety of experiences and prior learnings in the development of understanding of current problems or to attach greater meaning to present learnings. It would be indicated in the classroom by the telling of personal experiences to illustrate the problem or its solution, by the remarking or questioning upon certain similarities in the present situation to past lessons in the same or different subject-matter areas, by the attempts to apply the present learning to current problems outside the lesson itself, by the presence of evidence of the "aha" experience in students by voice or facial expression, etc. The absence of integration would be indicated by the absence of the above, and by the presence of such things as formal recitation of material studied, rote memory drills, etc. Essentially, this dimension should attempt to get at the formation in the students of habits of relating new experiences to old, of exploring new ideas to exhaust their meanings as opposed to the formation of the habits of acquiring isolated facts or dividing their experiences into "watertight compartments." This is perhaps related to later flexibility or rigidity in the meeting of problems, and to the building up of habits that will be useful in attempts to solve problems by the mechanism of repression.

Dependency. This dimension attempts to get at the extent to which the students are dependent on the teacher, the textbook, or other authority to determine the problems to be considered, the manner in which the problem is approached, and the solutions to be learned. It would seem to be the opposite of "initiative" or "self-direction" which, in the classroom context, would indicate the students' independence of any particular authority in settling the problems listed above. Essentially, this dimension seeks the answer to the question, who initiates the classroom activities -- the teacher or the students? Dependency would be indicated by student questions such as, "Just what do you want us to do?" and "How much do you want us to do?" It would also be indicated by dependence on a single textbook, by the questions being asked by the teacher rather than by the students, by frequent asking of the teacher, "Is this right?" This dimension is asking about the

*Acknowledgment is made of the assistance of Claire B. Ernhart and Wayne B. Holder, who were research assistants on this project.

extent to which the students are learning (by doing it) to rely on authority for the stimuli to solve problems, for the location of problems, and for the approval of solutions. Or are the students learning to have some confidence in their own ability to see, attack, and effectively solve problems on their own, and to evaluate the solutions they reach without having to go to an authority (teacher, book, etc.) to find out if they reached the "correct" solution. This dimension should be related to the student's future ability to think independently--that is, to see and formulate problems, to make appropriate investigations, and to evaluate proposed solutions.

Tension. This dimension refers to the concept of the world that the student is developing --is the world an unkind, hostile, fearful place, or is it a kind, friendly, secure sort of place. The dimension is primarily concerned with the presence of student reactions indicating fear and tension on the one hand and reactions indicating security and relaxation on the other. Indications of tension in the classroom would include student overreaction, student reluctance to perform, looking around to see where the teacher is, covert acts of disrespect for the teacher, covert fighting or teasing among the students. Indications of relaxation or security would include such things as a high degree of spontaneity in student actions, a confident approach of students to activities, performance of friendly and courteous acts toward the teacher and classmates. The dimension should indicate the extent to which the students are learning to show or experience fear and tension in social situations, as opposed to learning to relax and feel secure in such situations. This should be reflected in later life in their ability to work with other people in various situations, social, occupational, political, service, etc.

While these preliminary definitions served to focus attention on several aspects of significant classroom behaviors, it is almost immediately apparent that revisions may be needed as further work with them points out some of their limitations. For example, an inspection of the definitions suggests that Integration and Dependency may not be completely independent of each other. Such independence may or may not be possible, but from a measurement point of view it would be desirable. In other words, these are "armchair" definitions and, as such, should be subject to revision in the light of increased understanding obtained from future research and experience.

Development of the Observation Scales

A scale was developed for each of the three behavior dimensions by submitting possible items to a panel of judges and selecting those on which the judges agreed. Each item was in the form of a statement describing some kind of student behavior and was followed by five alternative responses indicating different amounts of frequencies, so that the observer had only to check one of the five responses to indicate how frequently the students engaged in that kind of behavior. For example:

1. Students erupt in emotional outbursts.
 - a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
4. Students recognize the teacher as the final arbiter on any question that arises.
 - a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
17. Students relate personal experiences to illustrate a problem or solution being discussed.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom

When a sufficient number of such items had been prepared for each dimension, they were assembled together, placed in random order, and mimeographed. This collection of items was then submitted to the panel of judges for assignment to one or more of the three dimensions. Accompanying the items were short definitions of the dimensions; the longer definitions given above were not used, since they included examples which were actually incorporated in the list of items. The definitions furnished the judges were as follows:

Integration is concerned with the extent to which the students organize and use a variety of experiences and prior learnings in the understanding of the problem under consideration and in attaching greater meaning to present learnings. Essentially, this involves the formation of habits of interrelating experiences as opposed to habits of sticking strictly to the materials at hand.

Dependency is concerned with behaviors which indicate reliance upon the teacher, the textbook, or other authority to determine the problems to be considered, the approach, and the solution. It is the opposite of self-direction, which would indicate independence of any particular authority in the solution of problems.

Tension refers to a feeling of hostility, insecurity, and fearfulness in the classroom. At one extreme are student reactions indicating fear and tension, and at the other extreme are student reactions indicating security and relaxation.

The judges were asked to "read each item carefully, evaluate it in terms of these three dimensions to determine in which of them it might function as a measuring device." They were also told that any item might be placed in none or in all of the categories. They were asked to indicate "the response category that should have the greatest weight in the direction indicated by the title of the dimension."

Twelve judges were used. All of them were teachers. Ten were teachers and practice teaching supervisors in the Laboratory School of the University of Missouri College of Education. The other two were instructors in the Department of Psychology (not included as raters in the try-out of the scales reported below). The decisions of the judges were tallied and the items were selected according to the following criteria:

1. The judges should agree on the assignment of the item to that dimension.
2. There should be a minimum of assignments of the item to more than one dimension.
3. About ten items should be selected for each of the three dimensions.

Table 1 shows the degree of the judges' agreement on the items that were finally selected. Ten items were finally selected for each dimension; no items were selected which would be scored on more than one dimension. The items selected for each dimension are given in Appendix A. These items, placed in a random order, then served as the three observation scales. (Appendix B.)

TABLE I
NUMBER OF ITEMS ASSIGNED TO THE THREE DIMENSIONS
AT DIFFERENT LEVELS OF AGREEMENT AMONG 12 JUDGES

Per Cent of Agreement	Number of Items Assigned to Dimension		
	Integration	Dependency	Tension
100	8	1	0
90 - 99	2	3	3
80 - 89	0	3	2
70 - 79	0	3	2
60 - 69	0	0	1
50 - 59	0	0	2

These scales were then used by observers in recording and reporting their observations of classrooms. Each of thirty classroom groups in the Psychology Department of the University of Missouri was observed by two trained observers. In all but one or two cases the observers visited the class together. The usual procedure was for the observers to enter the classroom with the students and to take seats in the rear of the room. During the class they would observe and take notes on student behavior. The observation period normally lasted for one class period, fifty minutes for most classes, one hundred minutes for some. The "Observer Check List" was filled out independently by the two observers shortly after the end of the observation period. Scoring was done later, using the a priori scoring weights as given in Appendix A.

Development of the Values Scales

One of the characteristics which was specified as desirable in the development and use of these scales was the definite provision for the application of values. Such a procedure seemed desirable in view of the likelihood that different educators would not agree on the specific values to be attached to different kinds of student behavior. If these procedures were to be useful to more than those who held the same values as the investigator, some procedures were necessary which would permit others to apply their own value judgments to the behaviors considered.

One approach that might have been used would have been to ask the educator to indicate the amount of integration, dependency, and tension he would like to see in a class. Some of the more sophisticated could have done this. But it would be possible for many others to use this device to pay their lip service to certain popular shibboleths. Other difficulties with this approach would include the different meanings attached to these labels by different persons, and the difficulties involved in specifying different degrees of integration, dependency, and tension. This approach, therefore, did not seem particularly to be desired.

Another approach, and the one used, was to have the individual specify his values in terms of the pupil behavior descriptions used in the observation of the classes. This would tend to avoid the arousal of stereotyped verbal behavior often seen as reactions to words such as integration, etc. It would also tend to bypass the problem of agreeing on the definition of the three dimensions, whose meaning then becomes a function of the items included in the scale. Finally, this procedure would permit the use of the scoring system developed for the observation scales in determining which part of the dimension the individual valued most highly. Use of this scoring system thus would permit direct comparison of values scores and observation scores.

It should be pointed out that this method does not completely avoid the problem of different meanings for different persons. Rather, it shifts this problem to the response categories ("how much is often"?). But here the problem can eventually be handled by item analysis and such scaling techniques as the method of reciprocal averages. Thus the problem was not completely handled but was transferred to a place where it could be dealt with in the future.

The scale of values developed, then, consisted of the items which had been previously developed for the use of classroom observers. To these items the person was asked to respond in a way which would give a description of his own "ideal" classroom. A copy of the scale appears in Appendix C. Scoring weights used were the same as those for the observation scales, making possible the direct comparison of "observation" scores and "values" scores.

The "Ideal College Classroom" check list was filled out by all of the eight raters before they were given the rating scales and while they were still relatively naive with respect to the plan of the investigation and the variables being studied.

The Rating Scales

A simple graphic rating scale was devised for each of the three dimensions. In constructing these scales an attempt was made to describe the extremes of each dimension in neutral or favorable terms; terms that might be used by an advocate of that kind of student behavior in describing the classroom he would like to see.

The rater was asked to consider the behavior of the students in the classroom and to evaluate the quality of the balance achieved between tendencies toward both extremes of the dimension. The judgment thus involved both the observation of the behavior exhibited in the class as well as the application of the rater's values to this behavior. The rater then recorded his judgment on a nine-point scale which ranged from "Ideal Balance" to "Poor Balance." A copy of the rating scales is presented in Appendix D.

Eight members of the Psychology faculty of the University of Missouri were used as the raters. This meant that the raters were rating the classes of each other as well as the classes taught by graduate students. In order to prevent any possible embarrassment from this arrangement, all reports were coded and identifying data removed before being examined or scored. Each of the thirty classes was assigned to one of these raters.

This procedure was not considered to be especially desirable but was dictated by limitations of time and budget. It meant that the demands on the time of the raters had to be held to an irreducible minimum, preventing the use of multiple ratings on the different classrooms. On the other hand, the interest of the faculty in learning more about their own teaching procedures, as well as the planned use of corrections for the values of each rater in treating the data, might be expected to ameliorate to some extent the limitations of this procedure.

Results

Reliability of observer scales. The type of reliability data that appeared to be most pertinent in the evaluation of the three scales was that indicating the relative consistency of the results on the same classes when used by different observers. This would be indicated by correlation coefficients between the scores obtained from the check lists of the two observers. Use of this method yielded a reliability coefficient of .81 for the integration scale, indicating fairly good agreement between observers when the type of scale and the small number of cases ($N = 30$) are considered. The reliability coefficient for the dependency scale was .77, just slightly less than that for the integration scale. The reliability coefficient for the tension scale was .50, indicating appreciably less agreement between the observers on this scale. However, all three of these coefficients were significant at the one per cent level.

Relations between observer scores, values scores, and ratings. The second aspect of this preliminary study of three proposed criterion measures was to observe certain characteristics of the proposed procedures. Specifically, it was desired to know whether or not the correction of the observer scores according to the values of the rater would yield results comparable to ratings which involved both observation and the application of values. That is, one might assume that a rater would assign the highest rating to a class which exhibited, say, "integration" behavior to the extent considered ideal by the rater. By the same token, classes in which the students' behavior differed from the rater's ideal should be rated lower. In other words, if a curve were plotted between the degree to which integrative behavior is exhibited by the members of classes and the ratings assigned to those classes, one would expect the curve to be curvilinear and more or less symmetrical, with a maximum point at the degree of "integration" which the rater considered ideal. Or if one were to plot the ratings against the absolute difference between the ideal and the observed degrees of integration, it should produce a curve which would be approximately linear.

These expectations were not borne out by the data of the present study. The correlation between ratings and the absolute difference between observer scores and rater's ideals on integration was .05. The corresponding correlation on the dependency dimension was -.19, and on the tension dimension it was -.29. Unfortunately, the data were insufficient to reveal the reasons for these results. Further analysis of the ratings by analysis of variance techniques indicated that the ratings were not significantly affected by the academic rank of the teachers nor by the differences in rank of the teacher and the rater. A significant difference was found between raters, but this was to be expected in a situation in which each rater rated different classes. Apparently the fault must be assumed, for the present, to lie either in the logic which led to these expectations or in the make-up of the scales themselves. If these difficulties were generated in the scales, there

are two possibilities that suggest themselves: (1) the particular wording in the rating scales may have suggested to the raters the consideration of behaviors different from those included in the dimensions as defined, or (2) since the scale for each dimension is rather short, it might be the case that the items selected may miss some important aspects of observable student behavior which most observers would consider to be indicative of the class's proper placement on the dimension.

Relationships between dimensions. As indicated earlier, it would be desirable to have dimensions that were independent of each other. Complete independence may or may not be possible, but a suspicion was expressed on the basis of the definitions of the dimensions that they might not be relatively independent. The limited data of this preliminary study indicated that this was indeed the case, as shown in Table 2.

TABLE 2
RELATIONSHIPS BETWEEN DIMENSIONS

	Dependency	Tension
Integration	-.89	-.58
Dependency		.51

Item analysis. In order to provide specific suggestions for future refinement of the scales, an item analysis was performed on the items of each scale. The results are given in Appendix A. The method used was the initial step in the reciprocal averages scaling technique, the computation of the mean total score associated with each possible response to an item. That is, on Item 2, the "Observer Check Lists" were sorted according to the response checked by the observer. This yielded four groups of papers. For each group, the mean total "integration score" was computed and entered in the column titled "Mean Total Score." A "good" item is indicated by a regular progression of these mean scores from one extreme of the possible responses to the other. According to this criterion, seven of the ten integration items appear to be operating more or less properly. The other three appear to need some attention. This was encouraging and may indicate that future work with a slightly modified integration scale might be quite productive.

Analysis of the items in the dependency scale was not so encouraging. There only four items could be considered to be functioning well. The remaining six items should be revised before further work is done with that scale.

The analysis of the tension items indicated that six of the items functioned fairly well, while four items appeared to require revision.

In considering the results of the item analysis, it should be noted that the limitations in the data, resulting from the size and nature of the sample of classrooms used, prevent one from taking the results at face value. That is, sampling only classes in psychology from a single institution might very well tend to utilize only a part of the range of variability that was built into the items. Thus for several items some of the extreme response categories were never used. This would normally be taken as evidence that these categories were not needed. But in considering the revision of items, some judgment (or some further research) would seem to be required in deciding whether or not these responses would be used in describing student behavior in classrooms in other fields or in other institutions.

Conclusions and Recommendations

It has been shown that it is possible to set up and use procedures which yield estimates of the degree to which the students in a class exhibit certain kinds of behavior. These estimates are fairly reliable in that two different users get results that are relatively consistent. Some validity

is assured by the use of competent judges in the selection of items. It has been suggested that such measures may be used to indicate the degree of success attained by the teacher in promoting student learning of certain kinds of behavior habits.

The attempt to use a combination of these estimates of classroom behavior and values attached to this behavior to predict a summary rating of the quality of the class was unsuccessful. The available data did not permit the assessment of possible reasons for this failure to produce the expected results. It was suggested that either (1) the rating scales were worded in such a way that the attention of the raters was drawn to different kinds of behavior than that specified in the definitions of the dimensions used, or (2) the limited sampling of behavior in each dimension may have resulted in the omission of certain behaviors that are important in specifying the proper placement of a class on that dimension.

Three possible behavior dimensions that might be measured in this way were defined and used as a basis for the measurements used in this study. These dimensions were concerned with the extent to which members of the class were practicing habits of integration, dependency, and tension. From the data on judges' agreement on the assignment of items to the dimension, the extent of agreement between observers in using the scale and the item analysis data, it would appear that a good beginning has been made on the development of a scale to measure integrative behavior of students in the classroom. Future development of this scale should, however, consider the possibility and desirability of extending the scale to provide more thorough coverage of other types of integrative behavior of students in a classroom.

The same types of data indicate that a promising start has been made in the development of measures of the degree of dependency being practiced by students in a classroom, although it appears to need more correction and further development than the integration scale.

The poorest showing among the three scales was made by the tension scale. Several items were assigned to this scale despite a relatively low degree of agreement among the judges. The reliability of this scale proved to be the lowest of the three scales. However, the item analysis indicates that there were several good items which may form the basis for the further improvement of the tension scale.

The future development of scales of this type for use as criterion measures in education should probably include an attempt to achieve a greater degree of independence of the scales, if such a development is possible. Some of the interrelationships observed between the scales may, of course, be the result of certain constellations of behavior in the classroom. This type of relationship could not be eliminated, nor should it be.

The present use of a priori scoring weights should be changed eventually to a more precise scoring system, such as that provided by the reciprocal averages method.

REFERENCES

1. ANDERSON, H. H., & BREWER, Helen M. Studies of teacher's classroom personalities, I. Dominative and socially integrative behavior of kindergarten teachers. Appl. Psychol. Monogr., 1945, No. 6.
2. ANDERSON, H. H., & BREWER, J. E. Studies of teacher's classroom personalities, II. Effects of teacher's dominative and integrative contacts on children's classroom behavior. Appl. Psychol. Monogr., 1946, No. 8.
3. ANDERSON, H. H., BREWER, J. E., & REED, M. F. Studies of teacher's classroom personalities, III. Follow-up studies of the effects of dominative and integrative contacts on children's behavior. Appl. Psychol. Monogr., 1946, No. 11.
4. ASH, F. E. The effect of teacher adjustment upon pupil adjustment. Univer. Iowa Doctoral Dissertations (1940-1941), 1944, 4, 76-79 (Abstract)
5. BARR, A. S., BECHDOLT, B. V., COXE, W. W., GAGE, N. L., ORLEANS, J. S., REMMERS, H. H., & RYANS, D. G. Report of the committee on the criteria of teacher effectiveness. Rev. educ. Res., 1952, 22, 238-263.
6. BOYNTON, P. L., DUGGER, H., & TURNER, M. The emotional stability of teachers and pupils. J. juv. Res., 1934, 18, 223-232.
7. BURGESS, T. C. Teacher behavior ratings as a criterion of teacher effectiveness. Technical report No. 1 for Contract Nonr 649(00), May, 1953.
8. GAGE, N. L. Recent developments significant for the criteria of counselor effectiveness. Paper read at Amer. Personnel & Guidance Assoc., Chicago, 1953.
9. MECHAM, G. P. A study of emotional instability of teachers and their pupils. George Peabody Coll. for Teachers, Contr. to Educ., 1941, No. 312. (Abstract)
10. ORLEANS, J. S., CLARKE, D., OSTREICHER, L., & STANDLEE, L. Some preliminary thoughts on the criteria of teacher effectiveness. J. educ. Res., 1952, 45, 641-648.
11. THRONKIKE, R. L. Personnel selection. New York: Wiley, 1949.
12. WHYTE, W. H., JR. Corporation and the wife. Fortune, Nov., 1951, 44, 109-111.
13. WHYTE, W. H., JR. Wives of management. Fortune, Oct., 1951, 44, 86-88.
14. WISPE, L. G. Evaluating section teaching methods in the introductory course. J. educ. Res., 1951, 45, 161-186.

APPENDIX A

ITEMS, SCORING WEIGHTS, AND ITEM ANALYSIS

Dimension: INTEGRATION

Question	Answer	Scoring Weight	Mean Total Score
2. Students bring up and discuss ideas that appear to be "original" within this group or context.	a. Often	4	----
	b. Fairly often	3	22.0
	c. Occasionally	2	17.8
	d. Once in a while	1	8.6
	e. Very seldom	0	3.3
6. Students remark on related problems that have not been considered.	a. Often	4	----
	b. Fairly often	3	19.5
	c. Occasionally	2	18.3
	d. Once in a while	1	9.0
	e. Very seldom	0	3.2
10. Students will discuss a problem in terms of what would happen if a given fact or event were not so or what would have happened if a given event had not happened.	a. Often	4	19.7
	b. Fairly often	3	8.0
	c. Occasionally	2	16.3
	d. Once in a while	1	17.4
	e. Very seldom	0	5.4
17. Students relate personal experiences to illustrate a problem or solution being discussed.	a. Often	4	----
	b. Fairly often	3	14.6
	c. Occasionally	2	18.0
	d. Once in a while	1	9.5
	e. Very seldom	0	6.2
19. Students criticize the ideas presented in their textbooks.	a. Often	4	25.5
	b. Fairly often	3	23.1
	c. Occasionally	2	20.9
	d. Once in a while	1	10.9
	e. Very seldom	0	5.7
20. Students inquire into the origins of a fact or an idea.	a. Often	4	25.5
	b. Fairly often	3	22.0
	c. Occasionally	2	18.3
	d. Once in a while	1	12.6
	e. Very seldom	0	4.5
24. Students explore relations of present topic to school topics presented in other contexts or other courses.	a. Often	4	24.5
	b. Fairly often	3	24.0
	c. Occasionally	2	20.5
	d. Once in a while	1	11.0
	e. Very seldom	0	5.0
27. Students discuss topics in reference to outside problems.	a. Often	4	17.7
	b. Fairly often	3	16.4
	c. Occasionally	2	15.0
	d. Once in a while	1	9.8
	e. Very seldom	0	2.9

Dimension: INTEGRATION

Question	Answer	Scoring Weight	Mean Total Score
28. Students appear to draw on many sources for their information.	a. A great deal	4	24.0
	b. Fairly much	3	22.8
	c. To some degree	2	14.3
	d. Comparatively little	1	6.2
	e. Not at all	0	2.3
29. Students explore relation of present topic to previous topics.	a. Often	4	13.5
	b. Fairly often	3	23.5
	c. Occasionally	2	16.2
	d. Once in a while	1	8.1
	e. Very seldom	0	4.0

Dimension: TENSION

Question	Answer	Scoring Weight	Mean Total Score
1. Students erupt in emotional outbursts.	a. A great deal	4	----
	b. Fairly much	3	7.8
	c. To some degree	2	9.2
	d. Comparatively little	1	7.4
	e. Not at all	0	8.3
3. Students look around to see where the teacher is.	a. Often	4	----
	b. Fairly often	3	----
	c. Occasionally	2	13.0
	d. Once in a while	1	9.6
	e. Very seldom	0	8.4
7. Students fight and/or tease each other in class.	a. A great deal	4	19.5
	b. Fairly much	3	11.5
	c. To some degree	2	10.9
	d. Comparatively little	1	9.1
	e. Not at all	0	7.9
8. Students blush, blanch, tremble, sweat, gulp or stammer when attention is directed to them.	a. Often	4	----
	b. Fairly often	3	----
	c. Occasionally	2	10.0
	d. Once in a while	1	8.6
	e. Very seldom	0	8.8
11. Students lower their eyes when their glance meets that of the teacher or the observer.	a. Always	4	----
	b. Often	3	9.5
	c. Occasionally	2	9.5
	d. Seldom	1	9.4
	e. Never	0	7.2

Dimension: TENSION

Question	Answer	Scoring Weight	Mean Total Score
12. Students drop papers, pencils, books, etc.	a. Often	4	----
	b. Fairly often	3	9.5
	c. Occasionally	2	13.0
	d. Once in a while	1	9.8
	e. Very seldom	0	7.8
15. Students mock the teacher surreptitiously.	a. Often	4	19.5
	b. Fairly often	3	----
	c. Occasionally	2	14.2
	d. Once in a while	1	10.4
	e. Very seldom	0	7.5
18. Students act worried.	a. A great deal	4	----
	b. Fairly much	3	8.7
	c. To some degree	2	10.6
	d. Comparatively little	1	8.4
	e. Not at all	0	7.6
22. Students engage in doodling, biting nails, playing with objects, fiddling, etc.	a. A great deal	4	14.0
	b. Fairly much	3	11.9
	c. To some degree	2	10.2
	d. Comparatively little	1	7.3
	e. Not at all	0	7.6
26. Free and comfortable laughter is heard in the classroom.	a. A great deal	0	----
	b. Fairly much	1	6.8
	c. To some degree	2	8.2
	d. Comparatively little	3	8.9
	e. Not at all	4	12.0

Dimension: DEPENDENCY

Question	Answer	Scoring Weight	Mean Total Score
4. Students recognize the teacher as the final arbiter on any question that arises.	a. Always	4	19.0
	b. Often	3	22.3
	c. Occasionally	2	16.0
	d. Seldom	1	11.7
	e. Never	0	9.0
5. Students appear satisfied if they answer teacher questions.	a. Always	4	25.5
	b. Often	3	23.0
	c. Occasionally	2	16.5
	d. Seldom	1	11.2
	e. Never	0	----

Dimension: DEPENDENCY

Question	Answer	Scoring Weight	Mean Total Score
9. Students appear satisfied to rely on what "the book" says.	a. Always	4	27.5
	b. Often	3	23.5
	c. Occasionally	2	19.6
	d. Seldom	1	13.6
	e. Never	0	7.5
13. Students are hesitant about committing themselves.	a. A great deal	4	26.7
	b. Fairly much	3	20.1
	c. To some degree	2	21.2
	d. Comparatively little	1	14.9
	e. Not at all	0	15.2
14. At the beginning of the period students wait for the teacher to start class activities.	a. Always	4	22.5
	b. Often	3	21.0
	c. Occasionally	2	28.5
	d. Seldom	1	----
	e. Never	0	----
16. Students ask teacher to specify the amount of work to be done.	a. Often	4	23.0
	b. Fairly often	3	24.5
	c. Occasionally	2	21.4
	d. Once in a while	1	10.4
	e. Very seldom	0	12.6
21. Students ask the teacher "Is this right?"	a. Often	4	21.6
	b. Fairly often	3	19.9
	c. Occasionally	2	20.5
	d. Once in a while	1	15.8
	e. Very seldom	0	12.5
23. Students ask the teacher to specify in detail what they are to do.	a. Often	4	22.5
	b. Fairly often	3	24.3
	c. Occasionally	2	22.0
	d. Once in a while	1	17.0
	e. Very seldom	0	13.5
25. Students' classroom comments sound "textbookish."	a. Often	4	----
	b. Fairly often	3	29.0
	c. Occasionally	2	19.9
	d. Once in a while	1	19.2
	e. Very seldom	0	16.3
30. Students test ideas by comparing them to "what the book says" to determine if the ideas are correct.	a. Often	4	22.5
	b. Fairly often	3	26.2
	c. Occasionally	2	21.5
	d. Once in a while	1	20.3
	e. Very seldom	0	16.3

APPENDIX B

Observer Check List

OBSERVER CHECK LIST

Observer_____ Class code_____

Teacher_____ Class_____

The Ideal College Classroom

This is a part of the Missouri Studies on Teacher Effectiveness being done under contract with the Office of Naval Research. The study of teacher effectiveness necessarily involves some consideration of teaching objectives which, in turn, require the making of certain value judgments. This questionnaire is an attempt to obtain information about the values used in evaluating the classroom behavior of college students.

We are asking for your assistance in specifying the kinds of student behavior that would be found in an ideal class in your own teaching field. You can help us by looking at each of the following items and indicating the amount or extent of this kind of behavior that would be found in an ideal classroom.

Sample: Students are attentive to what the teacher says

- a. Always
- b. Often
- c. Occasionally
- d. Once in a while
- e. Very seldom

(If you feel that students in the ideal class would always be attentive, you would encircle the "a.")

Your opinions will be kept confidential. You can help us in this by: (1) writing your name only in the space below, (2) making no unnecessary marks on the questionnaire, and (3) stapling the sheets together at the bottom before returning to us. Your name will be removed and your paper coded before anyone sees your responses.

Code _____

Your Name _____
(Please staple here)

1. Students erupt in emotional outbursts.
 - a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
2. Students bring up and discuss ideas that appear to be "original" within this group or context.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
3. Students look around to see where the teacher is.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
4. Students recognize the teacher as the final arbiter on any question that arises.
 - a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
5. Students appear satisfied if they answer teacher questions.
 - a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
6. Students remark on related problems that have not been considered.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom

7. Students fight and/or tease each other in class.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

8. Students blush, blanch, tremble, sweat, gulp or stammer when attention is directed to them.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

9. Students appear satisfied to rely on what "the book" says.

- a. Always
- b. Often
- c. Occasionally
- d. Seldom
- e. Never

10. Students will discuss a problem in terms of what would happen if a given fact or event were not so or what would have happened if a given event had not happened.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

11. Students lower their eyes when their glance meets that of the teacher or the observer.

- a. Always
- b. Often
- c. Occasionally
- d. Seldom
- e. Never

12. Students drop papers, pencils, books, etc.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

13. Students are hesitant about committing themselves.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

14. At the beginning of the period students wait for the teacher to start class activities.

- a. Always
- b. Often
- c. Occasionally
- d. Seldom
- e. Never

15. Students mock the teacher surreptitiously.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

16. Students ask teacher to specify the amount of work to be done.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

17. Students relate personal experiences to illustrate a problem or solution being discussed.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

18. Students act worried.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

19. Students criticise the ideas presented in their textbooks.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

20. Students inquire into the origins of a fact or an idea.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

21. Students ask the teacher "Is this right?"

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

22. Students engage in doodling, biting nails, playing with objects, fiddling, etc.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

23. Students ask the teacher to specify in detail what they are to do.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

24. Students explore relations of present topic to school topics presented in other contexts or other courses.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

25. Students' classroom comments sound "textbookish."

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

26. Free and comfortable laughter is heard in the classroom.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

27. Students discuss topics in reference to outside problems.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

28. Students appear to draw on many sources for their information.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

29. Students explore relation of present topic to previous topics.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

30. Students test ideas by comparing them to "what the book says" to determine if the ideas are correct.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

APPENDIX C

The Ideal College Classroom

1. Students erupt in emotional outbursts.
 - a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
2. Students bring up and discuss ideas that appear to be "original" within this group or context.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
3. Students look around to see where the teacher is.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
4. Students recognize the teacher as the final arbiter on any question that arises.
 - a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
5. Students appear satisfied if they answer teacher questions.
 - a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
6. Students remark on related problems that have not been considered.
 - a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom

7. Students fight and/or tease each other in class.
- a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
8. Students blush, blanch, tremble, sweat, gulp or stammer when attention is directed to them.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
9. Students appear satisfied to rely on what "the book" says.
- a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
10. Students will discuss a problem in terms of what would happen if a given fact or event were not so or what would have happened if a given event had not happened.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
11. Students lower their eyes when their glance meets that of the teacher or the observer.
- a. Always
 - b. Often
 - c. Occasionally
 - d. Seldom
 - e. Never
12. Students drop papers, pencils, books, etc.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom

13. Students are hesitant about committing themselves.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

14. At the beginning of the period students wait for the teacher to start class activities.

- a. Always
- b. Often
- c. Occasionally
- d. Seldom
- e. Never

15. Students mock the teacher surreptitiously.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

16. Students ask teacher to specify the amount of work to be done.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

17. Students relate personal experiences to illustrate a problem or solution being discussed.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

18. Students act worried.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

19. Students criticise the ideas presented in their textbooks.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

20. Students inquire into the origins of a fact or an idea.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

21. Students ask the teacher "Is this right?"

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

22. Students engage in doodling, biting nails, playing with objects, fiddling, etc.

- a. A great deal
- b. Fairly much
- c. To some degree
- d. Comparatively little
- e. Not at all

23. Students ask the teacher to specify in detail what they are to do.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

24. Students explore relations of present topic to school topics presented in other contexts or other courses.

- a. Often
- b. Fairly often
- c. Occasionally
- d. Once in a while
- e. Very seldom

25. Students' classroom comments sound "textbookish."
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
26. Free and comfortable laughter is heard in the classroom.
- a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
27. Students discuss topics in reference to outside problems.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
28. Students appear to draw on many sources for their information.
- a. A great deal
 - b. Fairly much
 - c. To some degree
 - d. Comparatively little
 - e. Not at all
29. Students explore relation of present topic to previous topics.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom
30. Students test ideas by comparing them to "what the book says" to determine if the ideas are correct.
- a. Often
 - b. Fairly often
 - c. Occasionally
 - d. Once in a while
 - e. Very seldom

APPENDIX D

Missouri Studies on Teaching Effectiveness, Classroom Rating Scale

Missouri Studies on Teaching Effectiveness

Classroom Rating Scale

Rater Code _____

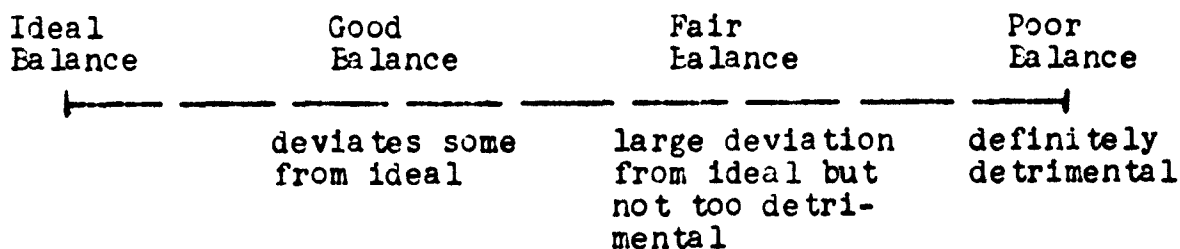
Class Code _____

Rater _____ Teacher _____ Class _____
(Please staple here)

Compare the observed classroom with your conception of the ideal classroom, i.e, that classroom which you feel would be most productive in terms of student development. We would like to have you rate the observed classroom on the extent to which it approaches the ideal on each of the following three dimensions.

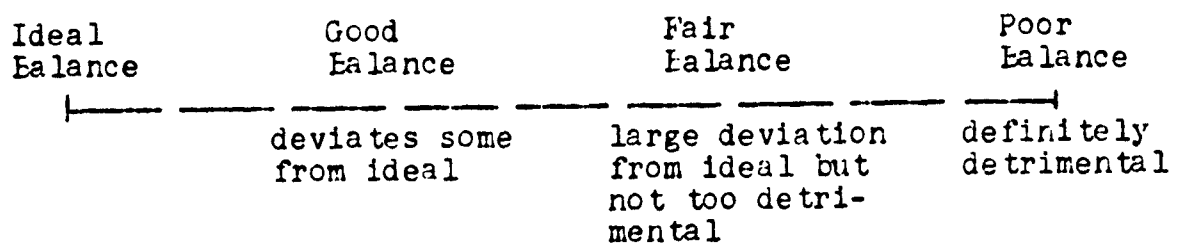
1. Consider the quality of the balance between "playing around with ideas" and "sticking to the facts." Did the teacher permit or produce a desirable amount of student activity in order to get them familiar with the ideas presented--to understand them and to relate them to other ideas and facts? Did the teacher draw the line before discussion went too far afield? Did the teacher set up a situation in which the students stayed on the topic and gave their attention to the facts and details of the course to a desirable degree?

Consider these things and rate on the quality of the balance achieved between these two directions of activity.



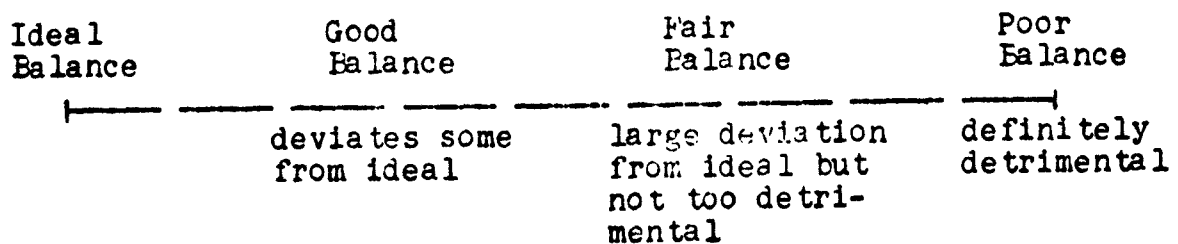
2. Consider the quality of the balance between the students' regard for the judgment and opinions of the teacher and textbook as opposed to the students' inclination to question or challenge the statements of the teacher or the textbook.

Rate on the quality of the balance achieved.



3. Consider the general state of student tension in the classroom. Are the students fearful, keyed-up, hostile? Or are they comfortable, relaxed, or "goofing off?" Or is there enough tension to keep them on their toes without any harmful effects.

Rate on the quality of the balance achieved.



APPENDIX E

DATA ON CLASSES OBSERVED

INTEGRATION

Class Code	Rater Code	Rater's Ideal	Observ. A	Observ. B	Observ. Mean	Obser- Values	Rating
9009	9383	39	1	0	.5	38.5	2
8814	4465	33	25	25	25.0	8.0	6
5496	9383	39	7	13	10.0	29.0	7
1670	2882	35	5	2	3.5	31.5	8
8082	0124	30	2	14	8.0	22.0	3
3971	4465	33	7	10	8.5	24.5	7
9912	0124	30	5	8	6.5	23.5	3
0030	9383	39	13	9	11.0	28.0	2
7253	3197	36	2	1	1.5	34.5	6
5810	2882	35	13	19	16.0	19.0	7
2720	7383	39	9	6	7.5	31.5	6
8160	0124	30	18	22	20.0	10.0	2
2312	8138	30	1	2	1.5	28.5	4
9481	0617	35	0	0	0.0	35.0	4
0797	5570	29	2	5	3.5	25.5	-
0960	9383	39	15	9	12.0	17.0	6
8834	0124	30	20	17	18.5	11.5	5
6071	8138	30	4	3	3.5	26.5	4
1266	3197	36	25	27	26.0	10.0	2
4792	0617	35	20	7	13.5	21.5	2
9632	5570	29	4	6	5.0	24.0	4
9937	8138	30	13	21	17.0	13.0	7
7768	8138	30	0	5	2.5	27.5	2
2023	4465	33	7	12	9.5	23.5	3
5261	0617	35	2	2	2.0	33.0	2
9993	5570	29	29	20	24.5	4.5	3
6555	5570	29	29	17	23.0	6.0	4
3613	2882	35	4	1	2.5	32.5	-
1098	2882	35	25	18	21.5	13.5	5
1033	0617	35	0	1	0.5	34.5	6

DEPENDENCY

Class Code	Rater Code	Rater's Ideal	Observ. A	Observ. B	Observ. Mean	Observ. Values	Rating
9009	9383	8	30	27	28.5	20.5	3
8814	4465	26	9	5	7.0	19.0	6
5496	9383	8	18	17	17.5	9.5	6
1670	2882	20	29	28	28.5	8.5	6
8082	0124	20	19	20	19.5	0.5	6
3971	4465	26	22	19	20.5	5.5	6
9912	0124	20	23	17	20.0	0.0	5
0030	9383	8	18	16	17.0	9.0	2
7253	3197	12	22	29	25.5	13.5	-
5810	2882	20	20	17	18.5	1.5	7
2720	9383	8	24	24	24.0	16.0	4
8160	0124	20	10	11	10.5	9.5	4
2312	8138	16	20	23	21.5	5.5	5
9481	0617	12	20	25	22.5	10.5	4
0797	5570	18	25	18	21.5	3.5	-
0960	9383	8	27	24	25.5	17.5	6
8834	0124	20	16	15	15.5	4.5	5
6071	8138	16	25	22	23.5	7.5	6
1266	3197	12	10	8	9.0	3.0	-
4792	0617	12	13	17	15.0	3.0	4
9632	5570	18	28	27	27.5	9.5	4
9937	8138	16	17	12	14.5	1.5	5
7668	8138	16	20	23	21.5	5.5	6
2023	4465	26	17	16	16.5	9.5	6
5261	0617	12	17	26	21.5	9.5	3
9993	5570	18	8	8	8.0	10.0	3
6555	5570	18	9	19	14.0	4.0	4
3613	2882	20	21	27	24.0	4.0	-
1098	2882	20	9	15	12.0	8.0	7
1033	0617	12	23	28	25.5	13.5	6

TENSION

Class Code	Rater Code	Rater's Ideal	Observ. A	Observ. B	Observ. Mean	Observ. Values	Rating
9009	9383	9	14	14	14.0	5.0	1
8814	4465	3	4	6	5.0	2.0	7
5496	9383	9	9	10	9.5	0.5	7
1670	2882	7	13	7	10.0	3.0	8
8082	0124	7	6	6	6.0	1.0	7
3971	4465	3	3	7	5.0	2.0	7
9912	0124	7	10	7	8.5	1.5	0
0030	9383	9	6	14	10.0	1.0	6
7253	3197	8	8	7	7.5	0.5	6
5810	2882	7	9	6	7.5	0.5	5
2720	9383	9	9	15	12.0	3.0	5
8160	0124	7	4	6	5.0	2.0	6
2312	8138	8	17	22	19.5	11.5	4
9481	0617	12	9	8	8.5	3.5	5
0797	5570	18	13	15	14.0	4.0	-
0960	9383	9	13	6	9.5	0.5	5
8834	0124	7	8	8	8.0	1.0	1
6071	8138	8	6	7	6.5	1.5	6
1266	3197	8	6	5	5.5	2.5	2
4792	0617	12	7	9	8.0	4.0	4
9632	5570	18	11	8	9.5	8.5	4
9937	8138	8	7	7	7.0	1.0	7
7668	8138	8	10	7	8.5	0.5	6
2023	4465	3	8	15	11.5	8.5	4
5261	0617	12	10	8	9.5	2.5	2
9993	5570	18	6	7	6.5	11.5	3
6555	5570	18	5	6	5.5	12.5	4
3613	2882	7	15	8	11.5	4.5	-
1098	2882	7	7	6	6.5	0.5	6
1033	0617	12	10	7	8.5	3.5	5